

Josh Akey, Associate Professor of Genome Sciences, University of Washington

Interviewer: And what's one question you studied recently in your, you like to talk about today?

Josh Akey: I guess we're going to do dogs, so one of our most recent studies has been in understanding the genetic differences between different breeds of dogs, dogs are an interesting models in genetics because they're so phenotypically diverse, so they look really different from one another and different breeds have very specific characteristics and so we're interested in those genetic basis of those characteristics

Interviewer: So if you would maybe talk about the cycle of describing or studying a particular question

Josh Akey: So I think the most important thing in science when figuring out what you want to study is to read a lot about the subject, to know a lot about everything that's been done and what gaps and knowledge exist, so that you can very precisely formulate the question that you want to ask and today usually the first step after you figuring out what question you want to ask, is to go to the computer and obtain data from databases, so one of the most kind of important changes that happened in genetics, in particular in science in general over the last decade has been that new technology has allowed the production of these massively large data sets and so all of that data gets deposited into databases that are freely available to anybody and so a lot of experiments these days are done by initially going to a computer and retrieving these large amounts of data and testing hypothesis based on data that already exists

Interviewer: Cool and then you move from using that to how do you add to the dataset then from those questions and hypotheses

Josh Akey: So I think there are two ways that you can add to the datasets, so the first is by figuring out a specific question you want to ask and retrieving data from databases and um and analyzing data and maybe producing more of that type of data for instance if you're interested in patterns of genetic variations between human populations and you're studying a particular gene you can obtain the sequence data that's been generated already for that gene and European populations and African populations and Asian populations but if you want to add to that you can generate new sequence data from a different African population or from a different European population and then contribute that data to these public databases and the second approach is that the data sets are so large and complex that, that there's a lot of information that is contained in them and that people haven't been able to extract yet and so when you develop new methods or you ask a new question you can use kind of this pre existing data to go and test those new hypothesis or to extract new information by using new approaches that haven't been applied to it yet, so I think those are two ways of kind of using databases in research today.